

Simulation and user analysis of BaBar data in a distributed cloud

We present a distributed cloud computing system that is being evaluated for the simulation and analysis of data from the BaBar experiment. The clouds include academic and commercial computing sites across Canada and the United States that are utilized in a unified infrastructure. Users retrieve a virtual machine (VM) with pre-installed application code; they modify the VM for their analysis and store them in a repository. The users prepare their job scripts as they would in a standard batch environment and submit them to a Condor job scheduler. The job scripts contain a link to the VM required for the job. A separate component, called Cloud Scheduler, reads the job queue and boots the required VM on one of the available compute clouds. The system is able to utilize clouds configured with various cloud Infrastructure-as-a-Service software such as Nimbus, Eucalyptus and Amazon EC2. We find that the analysis jobs are able to run with high efficiency even if the data is located at distant locations. We will show that the distributed cloud system is an effective environment for user analysis and Monte Carlo simulation.

Primary authors : Dr. SOBIE, Randall (University of Victoria)